

Optimization of Fuzzy Tsukamoto Membership Function using Genetic Algorithm to Determine the River Water

Qoirul Kotimah, Wayan Firdaus Mahmudy, Vivi Nur Wijayaningrum

Faculty of Computer Science, Brawijaya University, Indonesia

Article Info

Article history:

Received Oct 10, 2016

Revised Jun 18, 2017

Accepted Sept 11, 2017

Keywords:

Fuzzy tsukamoto

Genetic algorithm

Membership function

Water quality

ABSTRACT

Some aquatic ecosystems in rivers depend on the river water, so it needs to be maintained by measuring and analyzing the river water quality. STORET is one of the methods used to measure the river water quality, but it takes a quite high of time and costs. Fuzzy Tsukamoto is an alternative method that works by grouping the river water data, but it is difficult to determine the membership function value. The solution offered in this study is the use of genetic algorithm to determine the membership function value of each criterion. Based on the test results, the optimization of fuzzy membership function using genetic algorithm provides higher accuracy value that is 95%, while the accuracy value without optimization process is 90%. The parameters used in genetic algorithm are as follows: population size is 80, generation number is 175, crossover rate (*cr*) is 0.6, and mutation rate (*mr*) is 0.4.

Copyright © 2017 Institute of Advanced Engineering and Science.

All rights reserved.

Corresponding Author:

Vivi Nur Wijayaningrum,
Faculty of Computer Science,
Brawijaya University,
ROC Veteran Road, Malang, Indonesia.
Email: lsntl@ccu.edu.tw

1. INTRODUCTION

Water is one of the natural resources that is very important to human beings and other beings. The river is one of the place or container to hold the water resources flowing to the lowest point on the earth's surface. Rivers have a role as home to several aquatic ecosystems on earth that flows from upstream to downstream. The river is one of the nearest water sources for some residents in rural and urban areas in Indonesia. However, the river which an important part of human life sources have been polluted. It is mostly caused by human activity that occurred in the catchment area which became a river water supply and then flowed to residential areas. Increased industrial activity and household activity became one of the main causes of pollution that occurred in the watershed.

The disposal of industrial waste and household waste that does not do the screening process contributed to a decrease in the quality and quantity of the river water. Therefore, it needs an effort to maintain the quality, quantity, and continuity of the water so the ecosystems that exist therein remain balanced [1]. This is done by monitoring and controlling the water pollution regularly. An effort to monitor and control the water pollution is done by measuring and analyzing the water quality so that the factors affecting the pollution can be known, as stipulated in the Indonesian Government Regulation No. 82 of 2001. Based on the Ministry of the Living Environment of Indonesia No. 115 of Article 2 of 2003, the guidelines used to determine the status of the water quality is using the STORET method or Pollution Index Method. Determination of the water quality using STORET method still done manually so it took quite a long time and needs a laboratory cost which is quite expensive. The development of existing technologies can provide solutions to assist in the calculation and determination of the water quality to reduce the time and cost required.

One of the research relating to the determination of the river water quality has been conducted by Alam, Soebroto, and Dewi (2015) using the Fuzzy Tsukamoto method. In this study, the river water quality is calculated based on the criteria of chemical physics that have been stipulated in the Indonesian Government Regulation No. 82 of 2001. The parameters used in this study are TSS, BOD, DO, pH, phenols, oil, and grease. This study uses the 60 river water data to calculate its quality using the Fuzzy Tsukamoto method. Then, the calculation results are compared to the STORET method. Based on test results obtained, there are 6 different data between the calculations using the Fuzzy Tsukamoto method and the STORET method, so the accuracy value obtained is 90% [2]. An optimization process on the fuzzy membership function of each criterion that has been set is required to get a higher accuracy value.

Research on optimization using genetic algorithm has been successfully carried out by some previous researchers to solve different problems. The genetic algorithm has been applied to solve flexible job-shop scheduling problem [3], ship's route scheduling [4], and frozen food distribution [5]. Therefore, this research uses the genetic algorithm to optimize the membership function in the fuzzy Tsukamoto to measure the river water quality. The water quality will be divided into four classes based on some predetermined criteria. The solution offered in this research is forming a membership function in the fuzzy Tsukamoto to measure the river water quality to obtain higher accuracy value.

2. WATER

Water is part of the life of the earth's surface that has an important role. Water as one of the primary materials required to meet the needs of many people, even by all living beings. According to the Indonesian Government Regulation No. 82 of 2001, the water quality classification defined into four classes as follows:

- Class I: water that can be used for drinking and or other uses that require the water quality equal to that usability.
- Class II: water that can be used for infrastructure/water recreation facilities, freshwater fish farming, animal husbandry, water to irrigate crops or other uses that require the water quality equal to that usability.
- Class III: water that can be used for freshwater fish farming, animal husbandry, water to irrigate crops or other uses that require the water quality equal to that usability.
- Class IV: water that can be used to irrigate crops or other uses that require the water quality equal to that usability.

Based on the Ministry of the Living Environment of Indonesia No. 01 of 2010, the water pollution is defined as the entry of living beings, substances, energy or other components into the water caused by human activity. This can degrade the water quality up to a certain level, which causes the water cannot function according to its usefulness. The water pollution or commonly called liquid waste that pollutes the watershed can be divided into domestic waste and industrial waste.

3. PHYSICOCHEMICAL PARAMETERS

To find out the occurrence of water pollution, water pollutant parameter is used as an indicator to be able to conduct a prevention and control of pollution that occurred. This research uses the guidelines of the Ministry of the Living Environment of Indonesia No. 115 of 2003 regarding Guidelines for Determination of Water Quality Status as a standard of the water quality parameters. Whereas the water quality standard referring Indonesian Government Regulation No. 82 of 2001 regarding Water Quality Management and Pollution Control. Some parameters of physicochemical to test the quality of the river water are as follows: temperature, Total Dissolved Solid (TDS), Total Suspended Solid (TSS), degree of acidity (pH), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), phosphate, nitrate (NO₃), Ammonia-Nitrogen (NH₃-N), arsenic, cobalt, barium, boron, selenium, cadmium, chromium (VI), copper, iron, and lead.

STORET (storage and retrieval) is a method used to assess water quality status. The basic concept of STORET is comparing the data of water quality and its standard [6]. Based on the guidelines of the Ministry of the Living Environment of Indonesia No. 115 of 2003, there are 48 parameters of water quality status according to STORET value system are as follows: Total Dissolved Solid (TDS), water temperature, electrical conductivity (DHL), brightness, mercury, arsenic, barium, flourine, cadmium, chromium (VI), manganese, sodium, Nitrate-Nitrogen (NO₃-N), Nitrite-Nitrogen (NO₂-N), Ammonia-Nitrogen (NH₃-N), degree of acidity (pH), selenium, zinc, cyanide, sulfate, hydrogen sulfide, copper, lead, Residual Sodium Carbonate (RSC), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), fats, oils, phosphate, phenol, chlorine, boron, nickel, bicarbonate, carbon dioxide, salinity, Dissolved Oxygen (DO),

aldrin, dieldrin, chlordane, dichlorodiphenyltrichloroethane (DDT), detergent, lindane, Polychlorinated Biphenyls (PCB), endrine, benzene hexachloride (BHC), fecal coliform, and total coliform. Determination of water quality using STORET have to use all of these parameters while in Fuzzy Tsukamoto, the water quality can be calculated only with 8 parameters as follows: Total Suspended Solid (TSS), Biological Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Dissolved Oxygen (DO), degree of acidity (pH), phenol, fats, and oils. The use of Fuzzy Tsukamoto to determine the water quality by using fewer parameters than STORET will be able to save the cost of inspection.

4. FUZZY LOGIC

Lothfi A. Zadeh introduced fuzzy logic, which is the set theory of logic to overcome the concept of value between the truth values 'true' and 'false'. Fuzzy logic adopts the human way of thinking so that the value is not only 0 and 1, but also all the possibilities between 0 and 1 [7]. The language variables introduced to describe fuzzy phenomena in the natural language quantitatively. If the value is expressed in terms of language, fuzzy truth value to be very true, true, fairly true, less true, and not true [8].

In fuzzy concept, formulating an input to an output using fuzzy logic called fuzzy reasoning. Fuzzy reasoning can be solved using Tsukamoto, Mamdani, and Sugeno. Each type of reasoning will provide a different way to get the output. Fuzzy reasoning consists of five main parts as follows [9]:

- Fuzzification of input variables in the form of crisp data.
- The use of fuzzy operator (OR or AND).
- The implications of the premise.
- Tggregation effect based on the rule base that has been determined.
- Defuzzification.

5. GENETIC ALGORITHM

The genetic algorithm (GA) inspired by evolutionary biology, which used to find approximate solutions to an issue, especially in optimization problems [10]. GA has a population consisting of some chromosomes that represent possible solutions. GA has three main processes to form a new generation in each iteration that is crossover, mutation, and selection [11]. These processes are the basic principles of GA to conduct exploration and exploitation among the possible solutions [12].

The process of genetic algorithm as follows [13]:

- Generating a population consisting of several random individual chromosomes, that have the composition of specific genes.
- Calculating the fitness value of each individual.
- Conducting the reproduction process to produce offspring by performing crossover and mutation.
- Selecting individuals from a population consisting of parents and offspring to stay alive for the next generations to replace the old population.

6. METHODOLOGY

The data used in this study consists of 60 records of the river water sample. This data will be grouped into four classes according to the classification of the water quality that has been regulated by Indonesian Government Regulation No. 82 of 2001. The water quality symbolized by A for class I, B for class II, C for class III, and D for class IV.

6.1. Chromosome Representation

The first step in the GA process is determining the type of chromosome representation that will be used. In this research, the type of chromosome representation used is a real code. Each chromosome has genes as much as 20 which represents the value of the fuzzy membership functions for all criteria. Figure 1 shows a chromosome representation that can be formed.

gene	50	110	3	4	8	14	8	23	70	120	3	4	6	7	6	9	0.0012	0.0023	1.5	2
parameter	TSS		BOD				COD				DO			pH		phenol			fats and oils	

Figure 1. Chromosome representation

The gene values in each chromosome as shown in Figure 1 is used as boundary values of membership function on fuzzy. The membership functions for all parameters used are shown in Figure 2 to Figure 8.

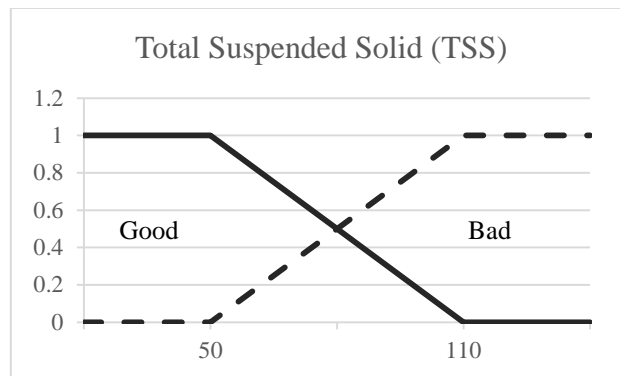


Figure 2. Membership function of TSS

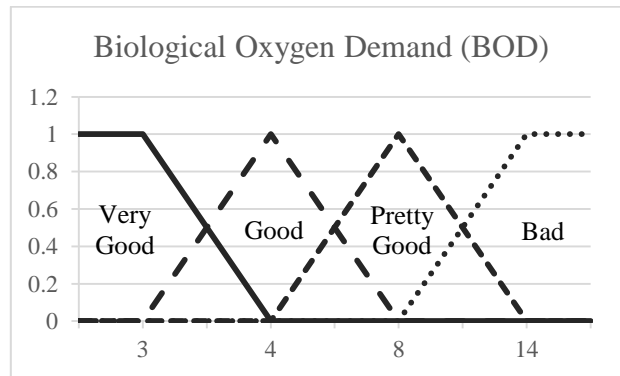


Figure 3. Membership function of BOD

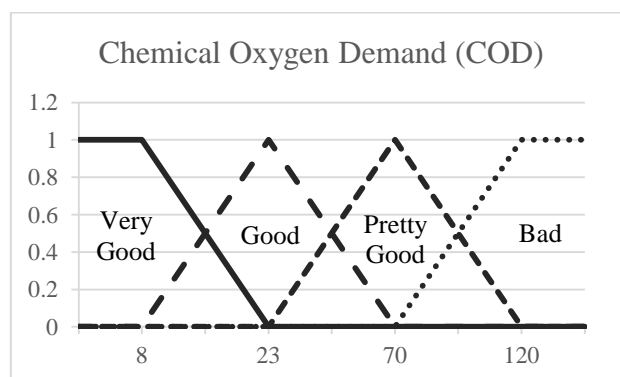


Figure 4. Membership function of COD

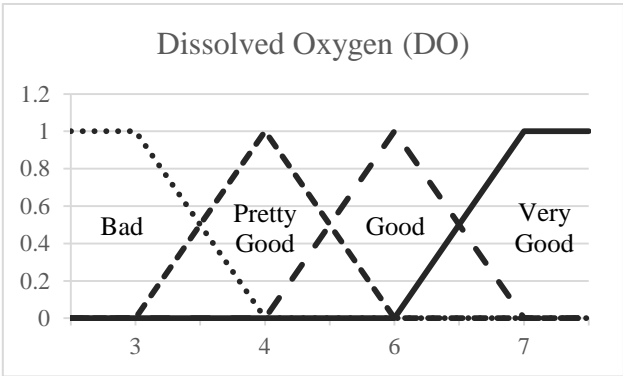


Figure 5. Membership function of DO

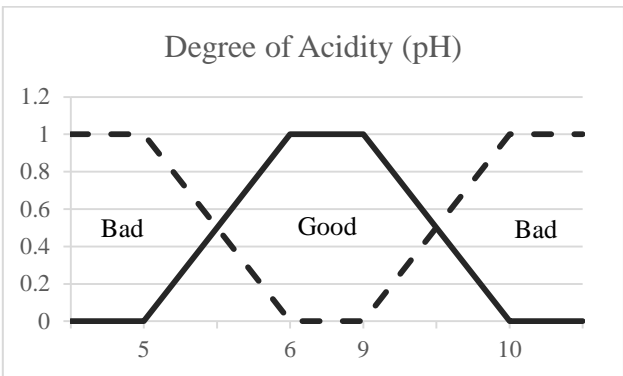


Figure 6. Membership function of pH

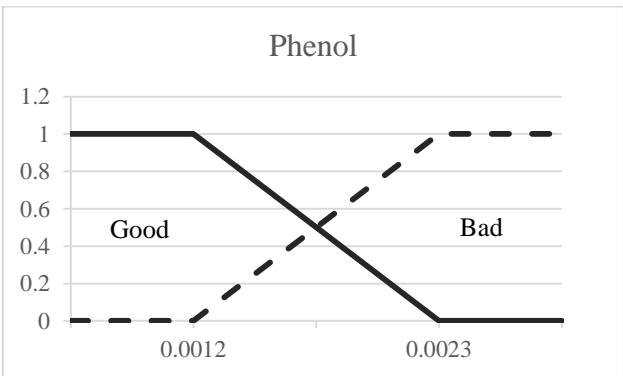


Figure 7. Membership function of phenol

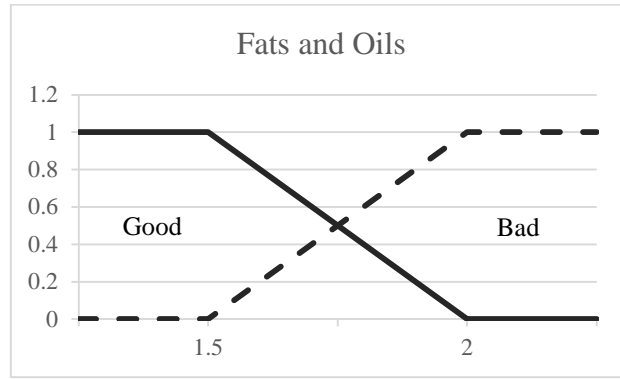


Figure 8. Membership function of fats and oils

6.2. Fitness Function

To calculate the fitness values, the gene values on each chromosome are applied to the fuzzy membership function according to the parameters that have been determined. Then, 60 records of the river water samples are predicted their class type using fuzzy Tsukamoto based on the membership function that has formed. Having obtained their class type, the accuracy value calculated by comparing the classification results using fuzzy Tsukamoto with the classification results using the STORET method. The fitness value is the accuracy value calculated using Equation (1).

$$\text{fitness} = \frac{\text{the number of an appropriate classification}}{\text{the number of data records}} \quad (1)$$

6.3. Crossover

The crossover process works by exchanging information between the chromosomes of two parents [14]. Crossover provides a method to explore a new region in the search space [15].

The type of crossover method used is extended intermediate crossover, which works by combining genes from both parents to produce two offspring [16]. The offspring formed using Equation 2, by choosing a random value in the range of values [-0.25, 1:25] for α .

$$\begin{aligned} \text{Offspring}_1 &= \text{Parent}_1 + \alpha(\text{Parent}_2 - \text{Parent}_1) \\ \text{Offspring}_2 &= \text{Parent}_2 + \alpha(\text{Parent}_1 - \text{Parent}_2) \end{aligned} \quad (2)$$

6.4. Mutation

The mutation process prevents the obtained solutions fall into local optimum because it provides a diversity of chromosomes in a population [14]. Mutation plays a role in restoring the lost genetic information as well as disturbs genetic information randomly to help the exploration process in the search space [17].

The type of mutation used is the random mutation, which works by changing the value of the selected gene with a small random number. Suppose variable x_i has a value range from \min_i to \max_i and offspring formed is $C=[x'_i, \dots, x'_n]$, then the offspring can be formed using Equation 3, by choosing a random value in the range of values [-0.1, 0.1] for r .

$$x'_i = x_i + r(\max_i - \min_i) \quad (3)$$

6.5. Selection

Selection is a step in GA which is used to select individuals that will be used in the next generation [18]. The selection method used is elitism. This selection types work by selecting individuals who have a greater fitness for survival.

All chromosomes in a population consisting of parents and offspring are combined into one and then sorted from the highest to the lowest fitness value. The top of chromosomes in the population as much as population size (popsize) taken for use in the next generation [19].

7. RESULTS AND DISCUSSION

Tests conducted in this study consist of three testing types, namely testing of the population size, testing of the combination of *cr* and *mr*, and testing the generation number. This testing is done to obtain the best parameters of the genetic algorithm. Each test scenario performed 10 times to obtain an average fitness value.

In the testing of population size, a population that is used is the numbers with multiples of 20 starting from number 20 to 200. The *cr* value used is 0.5 and the *mr* value is 0.5. The generation number used is 100. The test results of the population size shown in Figure 9.

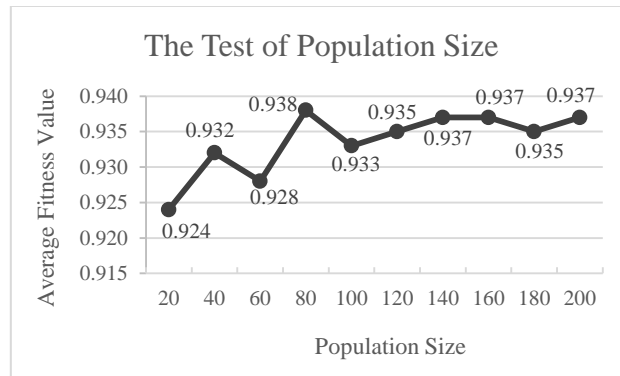


Figure 9. Test results of the population size

Figure 9 shows that the best average fitness value produced by population size of 80 is 0.938. The test results show that the bigger a population, the greater the possibility to obtain optimum solutions so that it can produce high fitness values.

The testing of the combination of *cr* and *mr* used to determine the most optimal *cr* and *mr* values in order to produce the best solution. The population size used is 80. The generation number used is 100. The test results of the combination of *cr* and *mr* are shown in Figure 10.

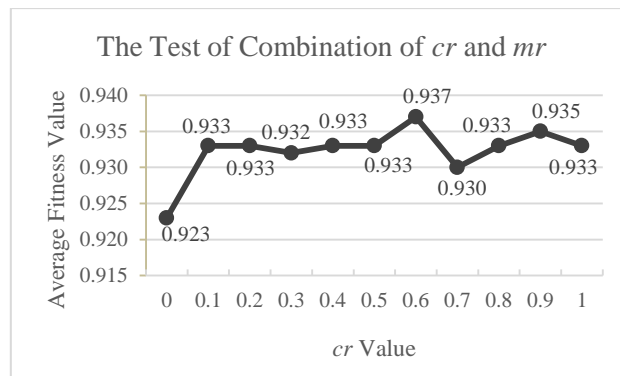


Figure 10. Test result of the combination of *cr* and *mr*

Figure 10 shows that the best average fitness value produced by the combination of *cr* is 0.6 and *mr* is 0.4. Based on Figure 10, it can be seen that the average fitness obtained is very diverse because of the absence of provisions *cr* and *mr* that should be used to obtain an optimal solution. Each issue will have a combined value of *cr* and *mr* that is different. If the value of *cr* and *mr* are not well defined, then the exploration and exploitation cannot produce a solution that is favorable, so premature convergence may occur [20].

The testing of generation number is done to get an optimal generation number to produce an optimal solution. The generation number used is the numbers with multiples of 25 starting from number 25 to 250.

The other parameters used are the parameters from the previous testing results as follows: the population size is 80, the cr value is 0.6, and the mr value is 0.4. The testing result of generation number is shown in Figure 11.

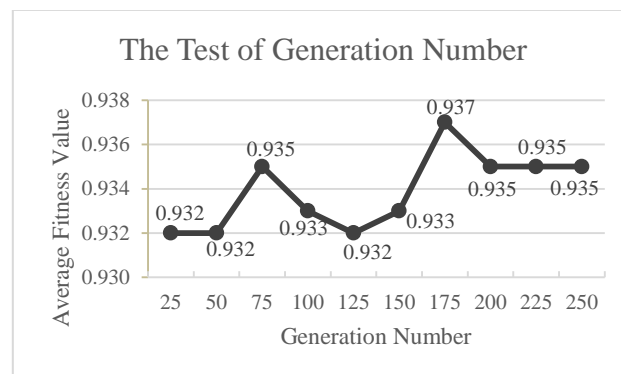


Figure 11. Test result of the generation number

Based on the test results of the generation number in Figure 11, the generation number as much as 175 produce the highest average fitness value. The generation number of 75 until 150 producing unstable number of fitness value. While the generation number of 175 until 250 has started to show stability in the absence of great change on the average fitness value. The more the generation number, the longer computation time required and the resulting solution is not necessarily optimal.

The parameters of the genetic algorithm which has been obtained from the testing results used to classify the 60 records of the river water sample. Table 1 shows a comparison of the water quality classification results in 2009 using STORET, Fuzzy Tsukamoto, and optimized Fuzzy Tsukamoto.

Table 2. Comparison of the water quality classification results in 2009

Months	Water Quality Classification Results		
	STORET	Fuzzy Tsukamoto	Optimized Fuzzy Tsukamoto
January	Moderately polluted	Moderately polluted	Moderately polluted
February	Heavily polluted	Heavily polluted	Heavily polluted
March	Moderately polluted	Moderately polluted	Moderately polluted
April	Moderately polluted	Moderately polluted	Moderately polluted
May	Moderately polluted	Moderately polluted	Moderately polluted
June	Moderately polluted	Moderately polluted	Moderately polluted
July	Lightly polluted	Lightly polluted	Moderately polluted
August	Moderately polluted	Moderately polluted	Moderately polluted
September	Moderately polluted	Moderately polluted	Moderately polluted
October	Heavily polluted	Moderately polluted	Heavily polluted
November	Moderately polluted	Moderately polluted	Moderately polluted
December	Moderately polluted	Moderately polluted	Moderately polluted

Based on the classification results using fuzzy Tsukamoto, 57 records of the river water sample produce the same class type of classification results were performed using STORET method. This indicates that the accuracy value obtained is 95%.

8. CONCLUSION

Based on the testing results and analysis that has been done, it can be concluded that the genetic algorithm can be used to optimize the membership function in the fuzzy Tsukamoto to determine the river water quality by using a real code representation. The genetic algorithm parameters that are used greatly affect the best chromosome that will be used as the value of membership function in the fuzzy Tsukamoto. The parameters to produce the best solution are as follows: population size is 80, generation number is 175, cr is 0.6, and mr is 0.4. The optimization process of membership function in the fuzzy Tsukamoto using genetic algorithm to measure the river water quality is proven to produce higher accuracy value that is equal to 95% compared to the measurement without optimization that is 90%.

In further research, the optimization of rule base in Fuzzy Inference System Tsukamoto can be done using the genetic algorithm so that the accuracy value obtained to be higher than the optimization process that is only done in the membership function. The use of hybridization of genetic algorithm and Variable Neighborhood Search (VNS) can also be developed on this issue so that the solution obtained optimal and efficient [21].

REFERENCES

- [1] Boskidis I, Gikas GD, Sylaios G, Tsihrintzis VA. "Water Quantity and Quality Assessment of Lower Nestos River," Greece. *J Environ Sci Heal Part A Toxic/Hazardous Subst Environ Eng*. 2011; 46(10):1050–67.
- [2] Alam GMR, Soebroto AA, Dewi C. "Implementasi Fuzzy Inference System (FIS) Metode Tsukamoto pada Sistem Pendukung Keputusan Penentuan Kualitas Air Sungai [Implementation of Fuzzy Inference System (FIS) Tsukamoto Method on Decision Support System for Determination of River Water Quality]." *DORO Repos J Mhs PTIIK Univ Brawijaya*. 2015; 5(6).
- [3] Mahmudy WF. "Solving Flexible Job-Shop Scheduling Problem Using Improved Real Coded Genetic Algorithms." *Int Conf Sci Technol Sustain*. 2014; 181–8.
- [4] Wijayaningrum VN, Mahmudy WF. "Optimization of Ship's Route Scheduling Using Genetic Algorithm." *Indones J Electr Eng Comput Sci*. 2016; 2(1):180–6.
- [5] Lesmawati W, Rahmi A, Mahmudy WF. "Optimization of Frozen Food Distribution using Genetic Algorithms." *J Environ Eng Sustain Technol*. 2016; 3(1):51–8.
- [6] Sholichin M, Othman F, Limantara LM. "Use of PI and STORET Methods to Evaluate Water Quality Status of Brantas River." *J Math Technol*. 2010; 3: 116–124.
- [7] Zadeh LA. *Fuzzy Sets*. *Inf Control*. 1965; 8: 338–53.
- [8] Zuojie W. "Concept Solution Generation for Product Innovation under User Fuzzy Semantic Requirements." *Int J Intell Eng Syst*. 2015; 8(3):1–10.
- [9] Mlakić D, Nikolovski SN, Knežević G. "An Adaptive Neuro-Fuzzy Inference System in Assessment of Technical Losses in Distribution Networks." *Int J Electr Comput Eng*. 2016; 6(3):1294–304.
- [10] Peng Z, Song B. "Research on Fault Diagnosis Method for Transformer based on Fuzzy Genetic Algorithm and Artificial Neural Network." *Kybernetes*. 2010; 39(8):1235–44.
- [11] Sivasankar S, Nair S, Judy MV. "Feature Reduction in Clinical Data Classification using Augmented Genetic Algorithm." *Int J Electr Comput Eng*. 2015; 5(6):1516–24.
- [12] Mahdi Homayouni S, Hong Tang S. "A Fuzzy Genetic Algorithm for Scheduling of Handling/Storage Equipment in Automated Container Terminals." *Int J Eng Technol*. 2015; 7(6):497–501.
- [13] B Wirawan Yohanes, H Handoko, H Kusuma Wardana, "Focused Crawler Optimization Using Genetic Algorithm," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 9, no. 3, pp. 403-410, 2011.
- [14] Lau HCW, Nakandala D, Zhao L. "Development of a Hybrid Fuzzy Genetic Algorithm Model for Solving Transportation Scheduling Problem." *J Inf Syst Technol Manag*. 2015; 12(3):505–24.
- [15] Coley DA. "An Introduction to Genetic Algorithms for Scientists and Engineers." World Scientific; 1999. 125-126 p.
- [16] Mühlenbein H, Schlierkamp-Voosen D. "Predictive Models for the Breeder Genetic Algorithm: Continuous Parameter Optimization." *Evol Comput*. 1993; 1:25–49.
- [17] Sivanandam SN, Deepa SN. "Introduction to Genetic Algorithms." New York, USA: Springer; 2008.
- [18] Singh S, Agrawal S. "Parameter Identification of the Glazed Photovoltaic Thermal System using Genetic Algorithm-Fuzzy System (GA-FS) Approach and its Comparative Study." *Energy Convers Manag*. 2015; 105: 763–71.
- [19] Mahmudy WF, Marian RM, Luong LHS. "Modeling and Optimization of Part Type Selection and Loading Problem in Flexible Manufacturing System Using Real Coded Genetic Algorithms." *Int J Electr Comput Electron Commun Eng*. 2013; 7(4):251–60.
- [20] Herrera F, Lozano M. "Fuzzy Adaptive Genetic Algorithms: Design, Taxonomy, and Future Directions." *Soft Comput*. 2003; 7(8):545–62.
- [21] Mahmudy WF, Marian RM, Luong LHS. "Hybrid Genetic Algorithms for Multi-Period Part Type Selection and Machine Loading Problems in Flexible Manufacturing System." In: *IEEE International Conference on Computational Intelligence and Cybernetics*. 2013. pp. 126–30.